The role of surface-based representations of shape in visual object recognition

Reppa, I.; Greville, W.J.; Leek, C.

Quarterly Journal of Experimental Psychology

DOI:
10.1080/17470218.2015.1014379

Published: 13/03/2015

THE ROLE OF SURFACE-BASED REPRESENTATIONS OF SHAPE IN VISUAL OBJECT RECOGNITION

Irene Reppa

Wales Institute for Cognitive Neuroscience,
Department of Psychology,
Swansea University, UK

W. James Greville

College of Medicine,
Swansea University, UK

&

E. Charles Leek

Wolfson Centre for Clinical and Cognitive Neuroscience,
School of Psychology,
Bangor University, UK

RUNNING HEAD: Surface-based primitives and shape representation

Please address any correspondence to Irene Reppa at i.reppa@swansea.ac.uk, Wales Institute for Cognitive Neuroscience, Department of Psychology, Swansea University, Swansea, UK, SA2 8PP; or to Charles Leek at e.c.leek@bangor.ac.uk, Wolfson Centre for Clinical and Cognitive Neuroscience, School of Psychology, Bangor University, Bangor, UK, LL57 2AS.

ABSTRACT

This study investigates the contribution of surface-based primitives to the perception of three-dimensional object shape. Observers matched subsets of closed contour fragments, surfaces and volumetric components to whole novel objects during a whole-part matching task. We manipulated two factors: Viewpoint (either same or different between component parts and whole objects) and target-distracter similarity. Similarity was varied in terms of systematic variation in non-accidental (NAP) or metric (MP) properties (NAPs) of individual parts. The results showed that whole-part matching was better for surface parts and volumes over closed contour fragments. However, there was no difference between surfaces and volumes even across changes in viewpoint between wholes and parts. The same pattern was found regardless of whether whole-part similarity varied by NAP or MP differences. The results provide new evidence supporting a role for surface-based image primitives in object shape perception.

<ins>Word count</ins>: 140

The human visual system is remarkably adept at recognising complex three-dimensional (3D) objects despite the variability in sensory information about object shape brought about changes in viewpoint, scale, translation and illumination. One of the fundamental issues that theories of object recognition must address is how the visual system represents object shape (e.g., Attneave, 1954; Biederman, 1987; Cristino, Conlan, Patterson & Leek, 2012; Davitt, Cristino, Wong & Leek, 2014; Edelman, 1999; Hummel & Biederman, 1992; Hummel & Stankiewicz, 1996; Leek, Cristino, Conlan, Patterson, Rodriguez & Johnston, 2012; Leek, Reppa, & Arguin, 2005; Leek, Reppa, Rodriguez, & Arguin, 2009; Marr & Nishihara, 1978; Pizlo, Sawada, Li, Kropatch & Steinman, 2010; Pizlo, 2008; Ullman, 2006).

In principle, object recognition could be supported by several different types of shape information at varying spatial scales. For example, some kinds of low-level image features can be computed at a relatively coarse spatial scale including edge co-linearity (parallelism), global elongation, principal axis curvature, symmetry, aspect ratio and scale. Such information may be sufficient for classifying objects under some conditions (e.g., differentiating a banana and a carrot on the basis of variation in principal axis curvature). In contrast, other image features may (and in some cases, must) be computed locally at finer spatial scales. These include contrasts in luminance used to determine edge boundaries and vertices as well as local variations in surface depth and curvature (e.g., Biederman, 1987; Cristino et al, 2012; Davitt, Cristino, Wong & Leek, et al, 2014; Leek et al, 2012; Marr & Nishihara, 1978; J.F. Norman, Todd, H.F. Norman, Clayton & McBridge, 2006). It has also been proposed that the visual system uses higher-level primitives to represent object shape (e.g., Barr, 1981; Bergevin & Levine, 1993; Biederman, 1987; Biederman & Cooper, 1991; Guzman, 1968; Krivic & Solina, 2004; Marr & Nishihara, 1978; Pentland, 1986; Ullman, Vidal-

Naquet & Sali, 2002; Zerroug & Nevatia, 1999), including volumetric parts, such as generalised cylinders (Brooks, 1981; Marr & Nishihara, 1978), super-quadrics (Barr, 1981; Pentland, 1986) and 3D geons (Biederman, 1985; 1987).

Here, we investigate evidence for another kind of primitive derived from edge-based approximations of object surface shape (Leek et al., 2005; 2009). There are several reasons why surfaces may be expected to play a key role in visual perception (e.g., Norman & Todd, 1996; Norman, Todd, & Phillips, 1995; J.F. Norman et al., 2006). For example, surfaces are important in constraining actions (digit placement) during prehensile movement (e.g., Servos, Goodale, & Jacobson, 1992), and have been shown to influence the distribution of object-based attention (Leek, Reppa & Tipper, 2003; Nakayama, He, & Shimojo, 1995; Nakayama & Shimojo, 1992; Reppa, Schmidt & Leek, 2012; Reppa & Leek, 2003; 2006). Surfaces are also important for the binding of object shape and other attributes such as colour, texture and shadow (Cate & Behrmann; 2010; Chainay, & Humphreys, 2001; Fan, Medioni, & Nevatia, 1989; Faugeras, 1984; Fisher, 1989; Leek et al., 2005; 2009; Marr & Nishihara, 1978). Therefore, surface information can be important in constraining attention and action.

The role of surfaces in object shape representation has been less widely examined, and in particular the question of whether surface-based primitives can contribute to the representation of object shape for recognition. This question was first investigated by Leek et al (2005) using a whole-part matching task. Observers viewed stimulus streams comprising a 3D novel object made of two distinct volumetric parts, and second ‘comparison part’ stimulus comprising a sub-set of the edge contour of the novel object. The comparison stimuli could contain a sub-set of contour fragments from the novel object, a non-volumetric configuration of spatially adjacent edge-defined surfaces, or one of the two volumetric parts. The task was to decide whether or not the comparison stimuli matched any sub-set of shape information in the whole novel object. The key finding was that while whole-part matching of surface and volumetric parts was faster than for contour fragments, there was no difference in performance between surfaces and volumes; that is, the configuration of surfaces into volumetric components afforded no matching advantage over non-volumetric configurations of surfaces.

Further evidence for the role of visible surfaces in object shape representation was reported by Leek et al (2009). In that study observers memorised a set of novel objects, each consisting of two adjacent volumetric parts. They then performed a primed recognition memory task in which they had to discriminate between studied and non-studied objects. Some primes contained only surfaces that were visible in the parent object (visible surface only primes), while other primes contained some surfaces which were occluded in the parent object via self-occlusion (occluded surface primes). All primes were masked and were either related or unrelated to the subsequent target object in terms matching shape information. The results showed robust priming effects for related vs. unrelated primes on subsequent recognition latency. Most importantly, however, priming magnitude was determined by the degree of correspondence between visible surfaces in the primes and targets. That is, priming decreased for primes containing some occluded surfaces relative to primes only containing visible surfaces. On the basis of these results, Leek et al (2005; 2009) argued that shape perception comprises a level of representation that specifies edge-based surface primitives – and that these primitives may be used to constrain image classification during object recognition (see Lee & Park, 2002; Fan, Medioni & Nevatia, 1989; Faugeras, 1984; Fisher, 1989 for related approaches in machine vision).

Current Study

The current evidence suggests that visible surfaces are explicit in object shape representation. Nevertheless, two issues may compromise the evidence for the of surface-based shape primitives in object shape representation. First, in the earlier behavioural studies based on whole-part matching by Leek et al (2005; 2009) whole objects and object parts were always presented from the same viewpoint. Viewpoint change may be critical in determining the kinds of shape representations that are computed from sensory input (e.g., Arguin & Leek, 2003; Foster & Gilson, 2002; Harris, Dux, Benito & Leek, 2006; Leek, 1998a; 1998b; Leek & Johnston, 2006; Leek, Atherton & Thierry, 2007; Tarr & Bülthoff, 1998; Ullman, 1998). Where there is no viewpoint change it may be sufficient for the visual system to compute image-based shape descriptions that can support direct matching of object shape components – whereas for the most part, object recognition systems must rely on the computation of shape representations that allow generalisation across views. For this reason, it is important to examine whether previous findings supporting the derivation of surface-based image primitives would generalise to tasks that require view generalisation.

A second limitation of current evidence is that previous studies have not systematically examined how variations in part similarity influence behavioural performance in whole-part matching. This is also of potential theoretical significance. A key distinction is between non-accidental (NAP) and metric (MP) properties of object shape (e.g., Biederman, 1987; Lowe, 1985). NAPs are elementary categorical dimensions that distinguish image features (e.g., straight vs. curved, parallel vs. tapered), and which are relatively invariant to viewpoint change. In contrast, MPs denote spatial relations among features that require precise specification (e.g., perceived aspect ratio, turning angle between contours and magnitude of curvature), and hence are sensitive to viewing angle. Results from several studies have shown that observers are more efficient at discriminating shape on the basis of changes in NAPs than MPs (e.g., Amir, Biederman & Hayworth, 2012; Biederman & Gerhardstein 1993; Biederman & Bar, 1999) – consistent with the hypothesis that the rapid computational of NAPs plays an important role in shape perception - and they are assumed to underlie the derivation of geon-based primitives in RBC (e.g., Biederman, 1987). Importantly, neither the study of Leek et al (2005) nor Leek et al (2009) investigated how variation between volumetric and surface parts in NAP/MP properties affected task performance between conditions. For instance in Leek et al (2005) match vs. mismatch decisions were made primarily on the basis of metric properties. That is, a mismatch volumetric or surface comparison part would come from an object that was similar to the target object in terms of non-accidental properties but differed mainly in term of metric properties. Therefore, any advantage in matching volumetric parts over surface parts may have been obscured because observers were forced to make 2D image-based discriminations. Thus, under conditions where the volumetric parts can be uniquely distinguished by variation in NAPs, then more efficient matching of volumetric relative to surface parts (contrary to the results of Leek et al, 2005) may be expected. Furthermore, the failure to control for NAP/MP distinctiveness may have given rise to a ‘spurious’ advantage in matching surface-based parts.

We examined these two issues in the current study using a whole-part matching paradigm similar to that used by Leek et al (2005). Observers were asked to make whole-part matching judgements to 3D multi-part novel objects. Part stimuli comprised regions of closed contour fragments, complete volumetric parts or edge-based surface polygons. Unlike in the Leek et al (2005) study, the parts were presented from either same viewpoint as shown in the whole object display, or from a different viewpoint. This manipulation allowed us to examine the generality of surface-based matching across viewpoint change.

In addition, we systematically varied the mismatch similarity between wholes and parts on mismatch trials in terms of NAPs or MPs. For one group of participants, mismatch parts were primarily distinguished from whole objects within a given trial by MPs (Figure 1A). In a second group of participants, mismatch comparison parts were primarily distinguished from the whole objects by NAPs. If the similarly efficient matching of surface and volumetric parts to whole objects was dependent on conditions where volumes can be distinguished only by MPs, then volumetric parts should outperform surface parts when the former can be uniquely distinguished by NAPs alone.

METHOD

Participants

Fifty participants were recruited from Swansea and Bangor Universities, twenty-five for each of the two mismatch groups. In the MP Mismatch Similarity group participants (7 males) had a mean age of 22.5 years (*SD*=3.21) and participated in the experiment for either course credit or £3 payment. In the NAP group participants (3 males) had a mean age of 21.5 years (*SD* = 5.13) and participated in the experiment for course credit. All participants reported normal or corrected-to-normal vision.

INSERT FIGURES 1A AND 1B ABOUT HERE

Apparatus and Stimuli

The experiment was run on a Windows PC with a 19" RGB monitor using E-Prime. The stimuli were twelve novel and geometrically regular 3D objects, each of which consisted of two connected volumetric parts: a main base or principal component, and a secondary component. They were rendered in externally lit, three-quarter views using Strata 3D Pro. Each object was scaled to fit within a 6 x 6-cm frame that subtended 6.86° x 6.86° of visual angle from a viewing distance of 50 cm. For each of the 12 objects, three types of comparison part stimuli were created: closed contour, volumetric, and surface parts (see Figure 1A and 1B). The volumetric parts comprised the principal (N=12) and secondary (N=12) components of each object. These differed in that when segmented from the 'parent' object the principal component contained a visible surface that was previously occluded when the components were combined. The secondary component contained only surfaces that were visible in both the whole (parent) object and individual part. For each object two types of surface parts were also created (N=24). These consisted of adjacent surfaces with the constraint that they did not make up a complete volume and that the number of surfaces matched exactly the number of surfaces in the volumetric parts stimuli. The contour stimuli (*N*=24) were created by selectively deleting regions of bounding and internal edge contour, so that the resulting closed form did not correspond to any single volume or any object surface. Following creation of the closed form, the surface information was removed by replacing the yellow colour with white (as the background).

Plane rotated versions of each part were created via rotation of +90^{o} or -90^{o} around the z axis perpendicular to the observer. These were used in the different viewpoint condition (see Design section and Figure 2).

INSERT FIGURE 2 ABOUT HERE

In order to prevent a strategy of simple pixel-by-pixel matching between comparison stimuli and whole objects in the same viewpoint condition, the whole object displays were enlarged to 150% the size of the images to be matched. Such moderate size transformations do not influence 3D-shape recognition (e.g., Fiser & Biederman, 1995; Norman et al., 2009). In addition, comparison part stimuli were centred so that the image pixels did not overlap.

As object and part stimuli necessarily differed in terms of low-level features (e.g. amount of visible contour and surface area; number of vertices and visible surfaces), Tables 1 and 2 formally quantify and compare, respectively, those differences on each of these dimensions in two-dimensional space. Their effect on matching performance is reported in the Results (footnote 1).

INSERT TABLE 1 AND 2 ABOUT HERE

Design

The experiment was based on a 3 (Part Type: contour, volume, surface); X 2 (Part Viewpoint: same vs. different); X 2 (Part Identity: principal vs. secondary) X Mismatch Similarity (NAPs vs. MP), with the latter factor manipulated between-participants. Mismatch Similarity was manipulated using two groups of stimuli that varied the similarity of target to mismatch parts in terms of NAPs and MP contrasts. In the MP group, the frequency of targets and mismatch volumetric parts differing solely by MPs relative to NAPs was 2:1. In the NAP group the ratio of MP to NAP differences was 1:2 – see Figures 1A and 1B. The reason for using different ratios as opposed to pure sets of MP or NAP mismatch pairs was to avoid the possibility that participants would tune selectively to MP or NAP differences biasing performance criteria. The manipulation of this factor allowed us to determine whether whole-part mismatch similarity in terms of NAP vs MPs contrasts could influence the pattern of results (see Introduction). The factor of Part Identity was included in light of the previous results of Leek et al (2009) showing that part-whole priming is influenced by the presence of previously occluded surfaces in segmented part primes. The principal components contained a surface that was occluded in the whole object, while the secondary components contained only visible surfaces. The factor of Part Viewpoint was manipulated to permit examination of the effects of viewpoint change between whole objects and parts on the patterns of performance across part types – addressing a limitation of previous work (see Introduction).

For each Mismatch Similarity participant group there were 144 match and 144 mismatch trials (N total = 288). There were 24 trials for each of the 12 within-subjects conditions. For each participant, each whole object was presented 24 times, and each part stimulus type was presented four times (twice in the same viewpoint as the whole object and twice in a different viewpoint). Trial order was randomized, within each of four blocks, for each participant. The dependent measure was D prime (*d'*) scores calculated using the hit and false alarm rate per condition.

Procedure

In match trials, the comparison part stimuli comprised a sub-set of shape information from the whole novel object that was presented in the same trial. In mismatch trials, the comparison part stimuli belonged to a different object. Stimulus pairings for the mismatch trials were determined by the mismatch similarity group. Trial procedure is shown in Figure 3. Participants were seated approximately 50cm from the monitor. Each trial began with the central presentation of a visual prompt '*Ready'?* until the participant initiated the trial sequence by pressing the space bar. Following a blank inter-stimulus interval of 750ms, one of the whole object stimuli appeared at screen centre for 1,200ms. Following a blank inter-stimulus interval of 750ms, a part stimulus was displayed in the centre of the screen until the participant made a response. Participants were informed that each part would be in the same orientation as the whole object preceding it, or plane-rotated clockwise (for half of the parts) or counter-clockwise (for the other half of the parts). The task was to decide as quickly and as accurately as possible whether or not the part stimulus came from the whole object that preceded it. Responses were made by pressing one of two keys (D or K) labelled *'Yes'* or *'No'* on a standard keyboard. There was a response deadline of 3 seconds. If a response was incorrect or timed out participants received feedback in the form of a short error tone. Half of the participants in each Mismatch Similarity group made match (Yes) responses with their dominant hand and mismatch (No) responses with their non-dominant hand. For the other half, these assignments were reversed. The experiment lasted approximately 35 minutes.

INSERT FIGURE 3 ABOUT HERE

RESULTS

Mean correct response times (RT) greater than or equal to 2 standard deviations (*SD)* from their own grand mean were excluded from the data. This accounted 0.9% of correct response RT. Timed-out responses accounted for .03% of all trials. The mean error rate across all conditions was 32.2% (*SD*=15.10%).

The goals of the analyses were to examine two key issues: (1) whether the patterns of whole-part matching across the contour, volume and surface part conditions is modulated by changes in viewpoint, and (2) whether the relative efficiency of matching volumetric vs. surface parts is dependent on the extent to which mismatch decisions are made primarily on the basis of NAP or MP differences.

Table 3 shows the mean *d'* per condition per Group. Mean *d'* scores per condition are shown in Figure 4.

INSERT TABLE 3 and FIGURE 4 ABOUT HERE

Part Viewpoint x Part Type x Part Identity. The goal of this analysis was to examine whether the pattern of matching for the three part types (contour, volumes, and surfaces) when they appear at the same viewpoint as the whole object (same viewpoint condition) changes when parts appear at a different viewpoint from the whole object (different viewpoint condition). The factor of Part Identity (principal vs. secondary) was included in the current analysis, because previous evidence (Leek et al., 2009; see Introduction) has shown significant differences in performance between comparison parts which contain occluded surfaces (as the current principal volumetric parts do) and parts which have only previously visible surfaces (the current secondary volumetric parts and the surface parts). For this analysis we collapsed across

Mismatch Similarity. Cell means per condition (collapsed across Mismatch Similarity) appear in Figure 4A.

A 2 (Part Viewpoint: same vs. different) x 3 (Part Type: contour, volumetric, surface) x 2 (Part Identity: principal vs. secondary) repeated-measures ANOVA showed a significant main effect of Part Viewpoint, F (1, 22) =43.76, p<=0001, with better performance with same versus different part viewpoint trials. There was also a significant main effect of Part Type, F (2, 44) =6.91, p=.002. The main effect of Part Identity was not significant, F (1, 22) =1.70, p>.05, but it was qualified by a significant Part Type and Part Identity interaction, F (2, 44) =3.96, p=.03. There were no other significant effects or interactions (all p values >.05).

Pairwise comparisons examining the Part Type X Part Identity interaction showed that for principal parts there was no difference between contour and volumetric parts, t (23)=.19, p>.05. However, surface parts yielded significantly higher sensitivity (*d'*) than contour and volumetric parts [(23)=2.58, p=.017, and t (23)=2.98, p=.007, respectively]. The pattern of *d'* was very different for secondary parts, where volumetric components contained no occluded surfaces. Here both volumetric and surface parts yielded higher *d'* scores than contour parts [t (23) =3.07, p=.005, and t (23) =2.45, p=.02, respectively]. Further comparisons between principal and secondary parts, showed no difference in *d'* between principal and secondary contour parts, t (23) =1.01, p>.05, and principal surface parts were only marginally more discriminable than their secondary versions, t (23)=1.93, p=.07. However, secondary volumetric parts yielded significantly higher *d'* than principal volumetric parts, t (23) =2.32, p=.02. This is further evidence of the performance cost associated with occluded surfaces, corroborating previous evidence by Leek et al (2009).

Group X Part Type (for secondary parts only). The second set of analyses examined whether the nature of similarity between comparison parts and whole objects (NAPs vs. MPs) might influence the pattern of matching. For this analysis we collapsed across Part Viewpoint, as it was not involved in an interaction with Part Type in the previous analysis. The *d'* scores from secondary parts only were used here. That is because for secondary parts, both volumetric and surface parts contained visible surfaces only. Therefore, all else being equal, we examined whether mismatch similarity influenced the pattern of matching across the three part types. Cell means (collapsed across Part Viewpoint) for this comparison appear in Figure 4B.

A mixed 2 (Mismatch Similarity: NAP vs. MP) X 3 (Part Type: contour, volume, surface) ANOVA carried out on *d'* scores showed a significant main effect of Part Type, F (2, 46) =4.43, p=.02. Pairwise t-tests showed no difference in *d'* between volumetric and surface parts, t (23) =.53, p>.05, while both part types had higher *d'* than contour parts [t (23)=3.07, p=.005, and t (23)=2.45, p=.02, respectively]. Neither the main effect of Mismatch Similarity (F (1, 23) =3.69, p=.07), nor the Mismatch Similarity X Part Type interaction were significant, F (2, 46)<1, p>.05[1].

[1] Because there were differences between comparison parts in terms of percent edge contour, number of vertices, and surface area we carried out three ANCOVAs on *d'* scores with each of those factors as covariates to examine whether the pattern of results was influenced by those differences. We collapsed across the variables of Part Viewpoint and Mismatch Similarity We created a new independent variable, Part, by combining the Part Type and Part Identity into a single factor with six levels: contour principal, contour secondary, volumetric principal, volumetric secondary, surface principal, and surface secondary. None of the three low-level features yielded either a significant main effect or was involved in an interaction with the Part factor. Furthermore, the pattern of *d'* cannot easily be explained by the pattern of those differences. For example, there are instances where there were significant differences in low level features between two conditions with no difference in accuracy and vice versa.

GENERAL DISCUSSION

Previously, it was shown that volumes do not necessarily enjoy an advantage in whole-part matching performance over lower-order shape descriptions of surfaces (e.g., Leek et al., 2005; 2009), leading to the hypothesis that surface-based shape primitives can mediate object shape representations of 3D objects. Aiming to provide further support to this hypothesis, the current study examined the two factors that may have obscured a benefit for volumetric parts or primes: part viewpoint relative to the whole object, and nature of the mismatch parts. Three main findings emerged. First, observers were better at matching volumetric and surface parts to whole objects than contour segments, replicating previous work (Leek et al., 2005; 2009). Most critically, observers were equally efficient at matching surface and volumetric parts regardless of whether whole objects and parts varied in viewpoint, or whether mismatch trial similarity was based primarily on NAP or MP differences. Finally, matching performance was modulated by whether part stimuli contained an occluded surface (principal vs. secondary parts).

Viewpoint change between the whole object and part influenced performance overall, with observers showing higher discriminability to parts that appeared at the same viewpoint as the whole object. This finding is consistent with previous demonstrations of how viewpoint may be critical in object recognition performance (e.g., Arguin & Leek, 2003; Foster & Gilson, 2002; Harris, Dux, Benito & Leek, 2006; Leek, 1998a; 1998b; Leek & Johnston, 2006; Leek, Atherton & Thierry, 2007; Tarr & Bülthoff, 1998; Ullman, 1998). Nevertheless, the pattern of matching performance was not influenced by part viewpoint in the current study. In whole-part matching sequences where there is no viewpoint change (e.g., Leek et al., 2005; 2009), matching can be possible by computing image-based shape descriptions, without the need to compute volumetric representations. The finding that part viewpoint change relative to the whole objects did not lead to such an advantage for volumes, suggests that the observed equivalent matching performance across surface and volumetric parts cannot be explained as an artefact of computing image-based descriptions, but reflected the computation of object shape.

Non-accidental properties of object shape (e.g., Biederman, 1987; Lowe, 1985) are known to play a key role in object shape perception, with observers being more efficient at discriminating shape on the basis of changes in NAPs than MPs (e.g., Amir, Biederman & Hayworth, 2012; Biederman & Gerhardstein 1993; Biederman & Bar, 1999). Matching decisions in the previous matching study by Leek et al., (2005) were largely made on the basis of metric properties, which may have forced observers to make image-based discriminations, obscuring any advantage of volumetric parts over non-volumetric surface configurations. In the current study, the nature of mismatch parts (whether they came from objects that differed from the target objects in terms of NAPs or in terms of MPs) did not influence overall sensitivity or the pattern of matching. Both volumetric and surface parts were matched more efficiently that contour parts, regardless of whether mismatch decisions were made on the basis of MPs or on the basis of NAPs. This suggests the benefit of parts containing visible surfaces (surface and volumetric parts) over comparison parts that do not (such as the contour parts) does not depend on whether matching is done on the basis of metric or non-accidental properties.

The effect of surface occlusion on performance – demonstrated by the modulation in whole-part matching between principal and secondary part stimuli, is consistent with the earlier report by Leek et al (2009). They found differential part-whole priming from primes containing only surfaces that are visible in the whole object versus primes containing a previously occluded surface (i.e., as in the principal components used in the current study). Their finding, along with the current results, provides further evidence for the existence of a level of shape representation in human vision that contains only visible (i.e., viewer-centred) descriptions of object surfaces, and provides more general support to other evidence implication surface-based representations in object perception (e.g., Cate & Behrmann; 2010; Chainay, & Humphreys, 2001; Fan, Medioni, & Nevatia, 1989; Faugeras, 1984; Fisher, 1989; Leek et al., 2005; 2009; Marr & Nishihara, 1978; Nakayama, He, & Shimojo, 1995; Nakayama & Shimojo, 1992).

One surface-based model of shape representation and recognition was outlined by Leek et al (2005 – see also Ashbrook, Fisher, Robertson & Werghi, 1998; Faugeras, 1984; Fisher, 1989, Lee & Park, 2002 for implementations of surface-based models in computer vision). According to this surface-based model, 3D objects are represented in terms of their constituent surface patches, whose shapes are approximated by 2D polygons (see also Phillips et al., 2003). Those closed polygons can be either regular or irregular in shape, a property which distinguishes them from NAP-based volumetric primitives (such as geons). Object surface configuration is encoded within a surface configuration map (see also Lee & Park, 2002). This map uses reference frames, defined by a three-dimensional spatial coordinate system, to specify the local pair-wise relations between spatially adjacent surfaces.

Of particular significance for the model is that the representations do not contain volumetric primitives, and the perception of object shape does not involve volumetric part decomposition as would be predicted by some accounts of object shape representation (e.g., Barr, 1981; Brooks, 1981; Marr & Nishihara, 1978; Pentland, 1986). This hypothesis can account for the current findings of equivalent performance in matching irrespective of whether the surfaces were in a volumetric or non-volumetric configuration. In Leek et al (2005), the encoding of higher-order structure is hypothesized to come via the development of correlational structure in the surface configuration map where local groups of spatially adjacent surfaces (i.e., from the same object region) form stronger patterns of inter-correlation than spatially separated surfaces (i.e., those from different regions of the object). Such emergent higher-order structure provides a basis for the links between lexical-semantic distinctions among object parts (e.g., feet, legs, bodies, arms, hands and heads) and object shape representations.

In conclusion, the current results extend those found in earlier studies using whole-part matching to show that efficient matching of part stimuli comprising spatial adjacent groups of edge-defined surface patches generalises across changes in viewpoint, and is not dependent on the whether mismatch similarity is defined primarily in terms of NAP or MP differences. Thus, the data provide further empirical support for the hypothesis that object representations in human vision underlying object recognition make use of surface primitives.

REFERENCES

Ashbrook, A.P., Fisher, R.B., Robertson, C. & Werghi, N. (1998). Finding surface correspondence for object recognition and registration using pairwise geometric histograms. *Proceedings of the 5th European Conference on Computer Vision, 2,* 674-680.

Ankrum, C., & Palmer, J. (1991). Memory for objects and parts. *Perception & Psychophysics, 50,*141–156.

Arguin, M., & Leek, E. C. (2003). Viewpoint invariance in visual object priming depends on prime-target asynchrony. *Perception & Psychophysics, 65,* 469–477.

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review, 61,* 183-193.

Barr, A.H. (1981). Superquadrics and angle-preserving transformations. *IEEE Computer Grapics and Applications, 1,* 1-11.

Bergevin, R. & Levine, M.D. (1993). Generic object recognition: Building and matching course descriptions from line drawings. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 15,* 19-63.

Biederman, I. (1985). Human image understanding: Recent research and a theory. *Computer Vision, Graphics and Image Processing, 32,* 29-73.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94,* 115–147.

Biederman, I., & Cooper, E. (1991). Priming contour-deleted images: Evidence for intermediate representations in visual object recognition. *Cognitive Psychology, 23,* 393–419.

Biederman, I., & Gerhardstein, P. C. (1993). Recognizing depth-rotated objects: Evidence and conditions for three-dimensional viewpoint invariance. *Journal of Experimental Psychology: Human Perception &Performance, 19,* 1162–1182.

Brooks, R. (1981). Symbolic reasoning among 3-dimensional and 2-dimensional images. *Artificial Intelligence, 17,* 285-349.

Bülthoff, H.H., & Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences of the United States of America, 89,* 60-64.

Cate, D. & Behrmann, M. (2010). Perceiving parts and shapes from concave surfaces. *Attention, Perception and Psychophysics, 72*, 153-167.

Chainay, H. & Humphreys, G.W. (2001). The real object advantage in agnosia: Evidence for a role of surface and depth information in object recognition. *Cognitive Neuropsychology, 18*, 175-191.

Cristino, F., Conlan, L., Patterson, L.C. & Leek, E.C. (2012). The appearance of shape in visual perception: Eye movement patterns during recognition and reaching. *Proceedings of the 3rd International Conference on Appearance*, Lulu Press. Edinburgh, 125-127.

Davitt, L., Cristino, F., Wong, A. & Leek, E.C. (in press). Fixation preference for concave surface discontinuities during object recognition generalises across levels of stimulus classification. *Journal of Experimental Psychology: Human Perception & Performance.*

Dickinson, S., Bergevin, R., Biederman, I., Eklundh, J.-O., Jain, A., Munck-Fairwood, R. & Pentland, A. (1997). Panel report: The potential of geons for generic 3-D object recognition. *Image and Vision Computing, 15*, 277-292.

Edelman, S. (1999). *Representation and recognition in vision.* Cambridge, MA: MIT Press.

Edelman, S. & Weinshall, D. (1991). A self-organizing multiple-view representation of 3-D objects. *Biological Cybernetics, 64*, 209-219.

Fan, T-J., Medioni, G. & Nevatia, R. (1989). Recognizing 3-D objects using surface descriptions. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 11*, 1140-1157.

Faugeras, O.D. (1984). New steps towards a flexible 3-D vision system for robotics. *Proceedings of the 7th International Conference on Pattern Recognition*, 796-805.

Fisher, R. (1989). *From surfaces to objects: Computer vision and three dimensional scene analysis*. Chichester. John Wiley and Sons.

Foster, D.H., & Gilson, S.J. (2002). Recognizing novel three-dimensional objects by summing signals from parts and views. *Proceedings of the Royal Society of London Series B-Biological Sciences, 269*, 1939-1947.

Grossberg, S., & Mingolla, E. (1985). Neural dynamics of perceptual grouping: Textures, boundaries and emergent segmentation. *Perception & Psychophysics, 38*, 141–171.

Guzman, A. (1968). Decomposition of a visual scene into three-dimensional bodies. *AFIPS Proceedings of the Fall Joint Computer Conference, 33*, 291-304.

Harris, I., Dux, P.E., Benito, C.T. & Leek, E.C. (2008). Orientation sensitivity at different stages of object processing: Evidence from repetition priming and naming. *PLoS ONE, 3 (5), e2256*

Hayward, W.G. & Tarr, M. J. (1997). Testing conditions for viewpoint invariance in object recognition. *Journal of Experimental Psychology: Human Perception & Performance, 23(5),* 1511-1521.

Hummel, J. E., & Stankiewicz, B. J. (1996). An architecture for rapid, hierarchical structural description. In T. Inui & J. McCelland (Eds.), *Attention and Performance XVI: On information integration in perception and communication* (pp.93-121). Cambridge, MA: MIT Press.

Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review*, *99,* 480–517.

Jackendoff, R. (1990). *Semantic structures*. MIT Press. Cambridge.

Koenderink, J.J., van Doorn, A.J. & Klappers, A.M.L. (1996). Pictorial surface attitude and local depth comparisons. *Perception & Psychophysics*, *58,* 163-173.

Krivic, J. & Solina, F. (2004). Part-level object recognition using superquadrics. *Computer Vision and Image Understanding*, *95*, 105-126.

Lee, Y.-L., & Park, R.-H. (2002). A surface-based approach to 3-D object recognition using a mean field annealing neural network. *Pattern Recognition*, *35,* 299–316.

Leek, E.C., Patterson, C., Paul, M.A., Rafal, R., & Cristino, F. (2012). Eye movements during object recognition in visual agnosia. *Neuropsychologia*, *50 (9),* 2142-2153.

Leek, E.C., Reppa, I., Rodiguez, E. & Arguin, M. (2009). Surface but not volumetric part structure mediates three-dimensional shape representation. *Quarterly Journal of Experimental Psychology*, *62*, 814-829.

Leek, E.C. Reppa, I. & Tipper, S. (2003). Inhibition-of-return for objects and locations in static displays. *Perception and Psychophysics, 65,* 388-395

Leek, E. C. (1998a). The analysis of viewpoint-dependent time costs in visual recognition. *Perception, 27,* 803–816.

Leek, E. C. (1998b). Effects of stimulus viewpoint on the identification of common polyoriented objects. *Psychonomic Bulletin & Review, 5*, 650–658.

Leek, E.C., Reppa, I. & Arguin, M. (2005). The structure of 3D object shape representations: Evidence from part-whole matching. *Journal of Experimental Psychology: Human Perception & Performance, 31*, 668-684.

Leek, E.C., Atherton, C.J. & Thierry, G. (2007). Computational mechanisms of object constancy for visual recognition revealed by event-related potentials. *Vision Research, 5*, 706-713.

Leek, E.C. & Johnston, S.J. (2006). A polarity effect in misoriented object recognition: The role of polar features in the computation of orientation-invariant shape representations. *Visual Cognition, 13*, 573-600.

Lo, T.W.R., Siebert, J.P. (2009). Local feature extraction and matching on range images: 2.5D SIFT. *Computer Vision and Image Understanding, 113*, 1235-1250.

Lowe, D.G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision, 60*, 91-110.

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London B: Biological Sciences, 200*, 269–294.

Mikolajczyk, K., Schmid, C. (2005). A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 27*, 1615-1630).

Nakayama, K, He, Z. & Shimojo, S. (1995). Visual surface representation: A critical link between lower-level and higher-level vision. In S.M. Kosslyn & D.

Osherson (Eds.), *Visual Cognition: Vol 2. An Invitation to Cognitive Science* (2nd ed, pp 1-70). Cambridge, MA: MIT Press.

Nakayama, K. & Shimojo, S. (1992). Experiencing and perceiving visual surfaces. *Science*, *257*, 1357-1363.

Parron, C. & Washburn, D. (2010). Contrasting the edge- and surface-based theories of object recognition: Behavioural evidence from Macaques (Macaca mulatta). *Journal of Experimental Psychology: Animal Behavior Processes, 36*, 148-157.

Pentland, A.P. (1986). Perceptual organization and the representation of natural form. *Artificial Intelligence, 28*, 293-331.

Phillips, F., Todd, J.T., Koenderink, J.J. & Kappers, A.M.L. (2003). Perceptual representation of visible surfaces. *Perception & Psychophysics, 65*, 747-762.

Pizlo, Z. (2008). *3D Shape: Its unique place in visual perception*. London: MIT Press.

Pizlo, Z., Sawada, T., Li, Y., Kropatch, W.G., & Steinman, R.M. (2010). New approach to the perception of 3D shape based on veridicality, complexity, symmetry and volume. *Vision Research, 50,* 1-11.

Reppa, I. & Leek, E.C. (2006). Structure-based modulation of IOR is triggered by object internal but not occluding shape features. *Quarterly Journal of Experimental Psychology, 59,* 1857-1866.

Reppa, I. & Leek, E.C. (2003). The modulation of inhibition-of-return across object-internal structure: Implications for theories of object-based attention. *Psychonomic Bulletin & Review, 10,* 493-50.

Reppa, I., Schmidt, W. & Leek, E.C. (2011). Successes and failures in obtaining attentional object-based cueing effects. *Attention, Perception & Psychophysics, 74* (1), 43-69

Riesenhuber, M. & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2*, 1019-1025.

Tarr, M. J., & Bulthoff, H. H. (1998). *Object recognition in man, monkey, and machine.* Cambridge, MA: MIT Press.

Ullman, S. (2006). Object recognition and segmentation by a fragment-based hierarchy. *Trends in Cognitive Sciences, 11,* 58-64.

Ullman, S., Vidal-Naquet, M. & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nature Neuroscience, 5*, 682-687.

Ullman, S. (1998) Three-dimensional object recognition based on the combination of views. In Tarr, M.J. & Bülthoff, H.H. (Eds) *Object recognition in man, monkey and machine*. Cambridge, MA: MIT Press (pp 21-44).

Zerroug, M. & Nevatia, R. (1999). Part-based 3D descriptions of complex objects from a single image. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 21*, 835-848.

Table 1: Image properties of the contour, volumetric and surface comparison part stimuli.

Part Identity	Part Type	Percent (%) of total edge contour		*N* vertices		Surface area	
		M	*SD*	M	*SD*	M	*SD*
Large	Contour	58.61	*8.27*	9.00	*1.04*	25.95	*7.28*
	Volumetric	68.68	*8.28*	10.08	*1.31*	22.51	*7.96*
	Surface	74.75	*11.53*	9.54	*1.15*	24.03	*5.41*
Small	Contour	51.49	*9.35*	7.83	*1.11*	15.52	*4.29*
	Volumetric	42.86	*13.68*	5.75	*1.42*	11.38	*4.42*
	Surface	64.80	*12.22*	9.00	*1.35*	16.32	*3.47*

Table 2: Comparisons between each of the three comparison parts along each of three types of low-level feature. Comparisons are reported separately for each Part Identity (large vs. small). Number of edge vertices refers to the total number of edge vertices (Y, T, and L) per comparison part stimulus. Surface area (calculated in centimetres using the ImageJ software, version 1.43) refers to the area enclosed by the bounding contour. Asterisks (*) follow each significant difference.

Part Identity	Comparison	Low-level feature	*t* Statistic
Large	Contour vs. Volume	Percent (%) of total edge contour	t (11)=2.98, p=.007*
		N vertices	t (11)=4.52, p=.0001*
		Surface area	t (11)=1.1, p>.05
	Contour vs. Surface	Percent (%) of total edge contour	t (11)=3.93, p=.001*
		N vertices	t (11)=2.38, p=.04*
		Surface area	t (11)=.73, p>.05
	Volume vs. Surface	Percent (%) of total edge contour	t (11)=1.48, p>.05
		N vertices	t (11)=6.27, p=.0001*
		Surface area	t (11)= .45, p>.05
Small	Contour vs. Volume	Percent (%) of total edge contour	t (11) = 1.80, p>.05
		N vertices	t (11) = 3.99, p=.001*
		Surface area	t (11) = 2.33, p=.03*
	Contour vs. Surface	Percent (%) of total edge contour	t (11)=2.99, p=.007*
		N vertices	t (11)=2.31, p=.03*
		Surface area	t (11)= .50, p>.05
	Volume vs. Surface	Percent (%) of total edge contour	t (11)=4.14, p=.0001*
		N vertices	t (11)=5.74, p=.0001*
		Surface area	t (11)=3.04, p=.006*

Table 3: Mean *d'* scores (standard deviations in parenthesis) per condition for the MPs group (match/mismatch decisions made on the basis of metric differences) and the NAPs (match/mismatch decisions made on the basis of NAP differences).

Part Viewpoint			MP group	NAP group	
	Part Type	Part Identity	*Mean d' (SD)*	*Mean d' (SD)*	
Same	Contour	Principal			
		Secondary			
	Volumetric	Principal			
		Secondary			
	Surface	Principal			
		Secondary			
Different	Contour	Principal			
		Secondary			
	Volumetric	Principal			
		Secondary			
	Surface	Principal			
		Secondary			

FIGURE LEGENDS

Figure 1A and 1B. The stimulus sets used for participants in the MP Mismatch Similarity group (1A) and participants in the NAP Mismatch Similarity group (1B). All the part stimuli are shown for each of the twelve objects. For each object its pair appears directly across from it. The NAP/MP column reports the type of difference between each volume of each object when compared with one of the two volumes of its paired object. For instance, in Figure 1A there is an MP difference between the top volume of Object 1 and the top volume of Object 7. Similarly, in Figure 1B there is a NAP difference between the top volume of Object 2 (pyramid) and the lower volume of Object 7 (truncated pyramid). The 'Type of difference' column shows the type of MP or NAP differences between the left side and the right side volumes for each object. In some cases where the difference is in terms of NAP, there is more than one difference between the volumes (e.g., in Figure 1B, the volumes differ both in terms the axis shape and in terms of their ending). Note: CS stands for cross section, and AS stands for aspect ratio.

Figure 2. An illustration of the contrasting displays used for the same and different Part viewpoint conditions across Part types.

Figure 3. The trial procedure.

Figure 4. The mean *d'* per Part Type and Part Identity, collapsed across Group (Naps vs. MP) and Part Viewpoint (same vs. different). Bars indicate standard error.

FIGURE 1(a)

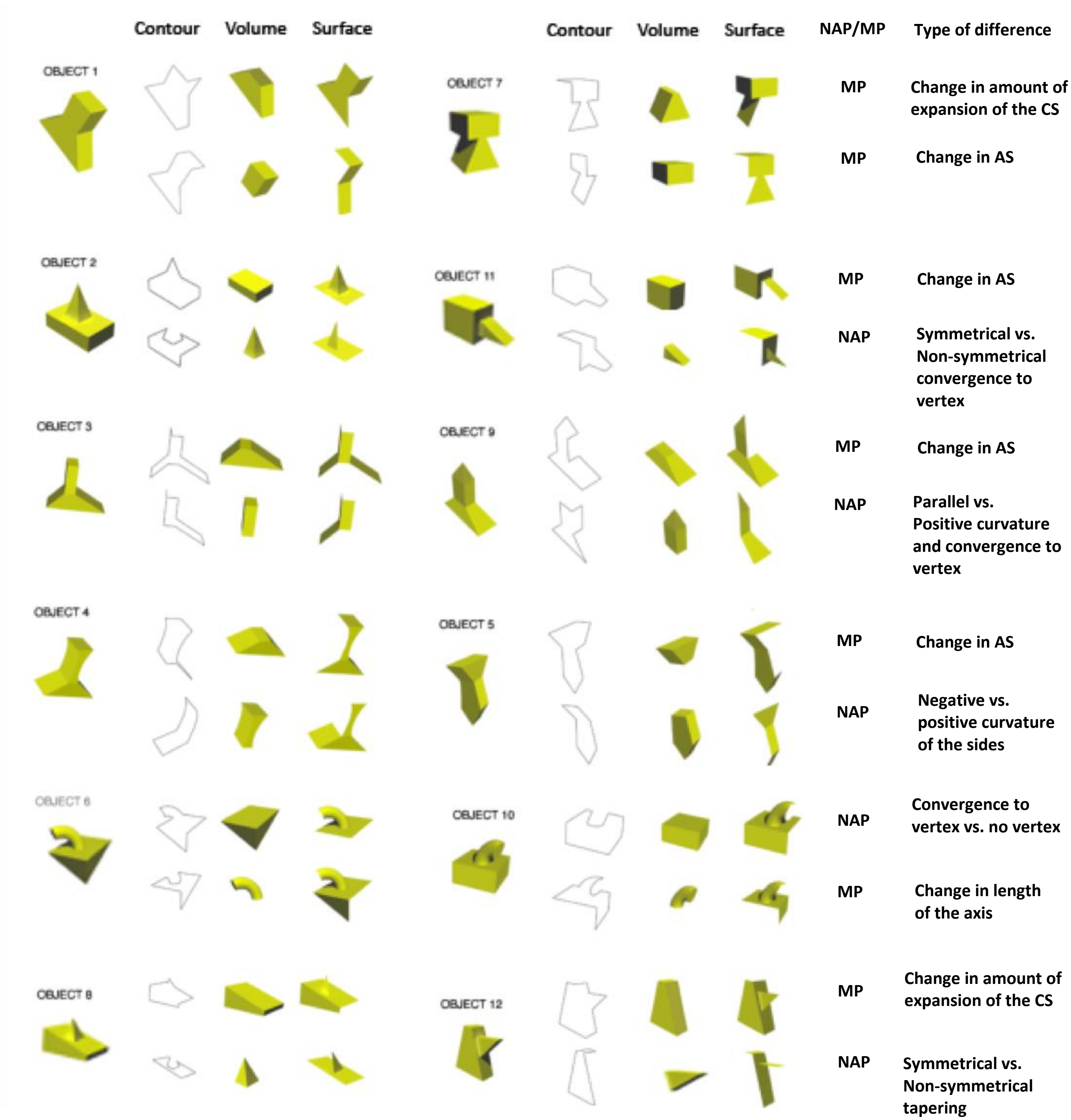
Contour
Volume
Surface
Contour
Volume
Surface
NAP/MP
Type of difference
OBJECT 1
OBJECT 7
MP
Change in amount of expansion of the CS
MP
Change in AS
OBJECT 2
OBJECT 11
MP
Change in AS
NAP
Symmetrical vs. Non-symmetrical convergence to vertex
OBJECT 3
OBJECT 9
MP
Change in AS
NAP
Parallel vs. Positive curvature and convergence to vertex
OBJECT 4
OBJECT 5
MP
Change in AS
NAP
Negative vs. positive curvature of the sides
OBJECT 6
OBJECT 10
NAP
Convergence to vertex vs. no vertex
MP
Change in length of the axis
OBJECT 8
OBJECT 12
MP
Change in amount of expansion of the CS
NAP
Symmetrical vs. Non-symmetrical tapering

FIGURE 1(b)

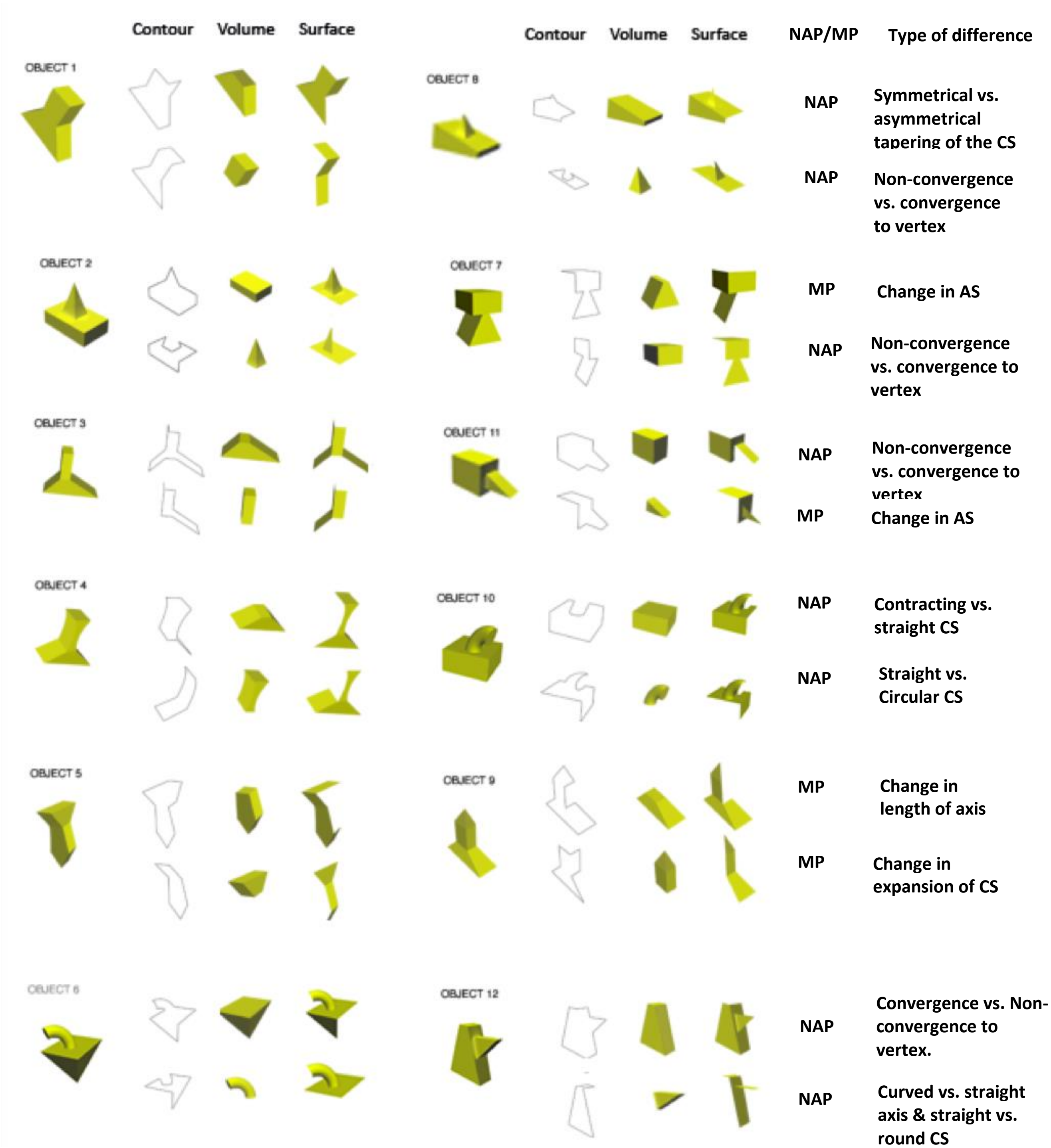
Contour
Volume
Surface
NAP/MP
Type of difference
OBJECT 1
OBJECT 2
OBJECT 3
OBJECT 4
OBJECT 5
OBJECT 6
OBJECT 8
NAP
Symmetrical vs. asymmetrical tapering of the CS
NAP
Non-convergence vs. convergence to vertex
OBJECT 7
MP
Change in AS
NAP
Non-convergence vs. convergence to vertex
OBJECT 11
NAP
Non-convergence vs. convergence to vertex
MP
Change in AS
OBJECT 10
NAP
Contracting vs. straight CS
NAP
Straight vs. Circular CS
OBJECT 9
MP
Change in length of axis
MP
Change in expansion of CS
OBJECT 12
NAP
Convergence vs. Non-convergence to vertex.
NAP
Curved vs. straight axis & straight vs. round CS

FIGURE 2

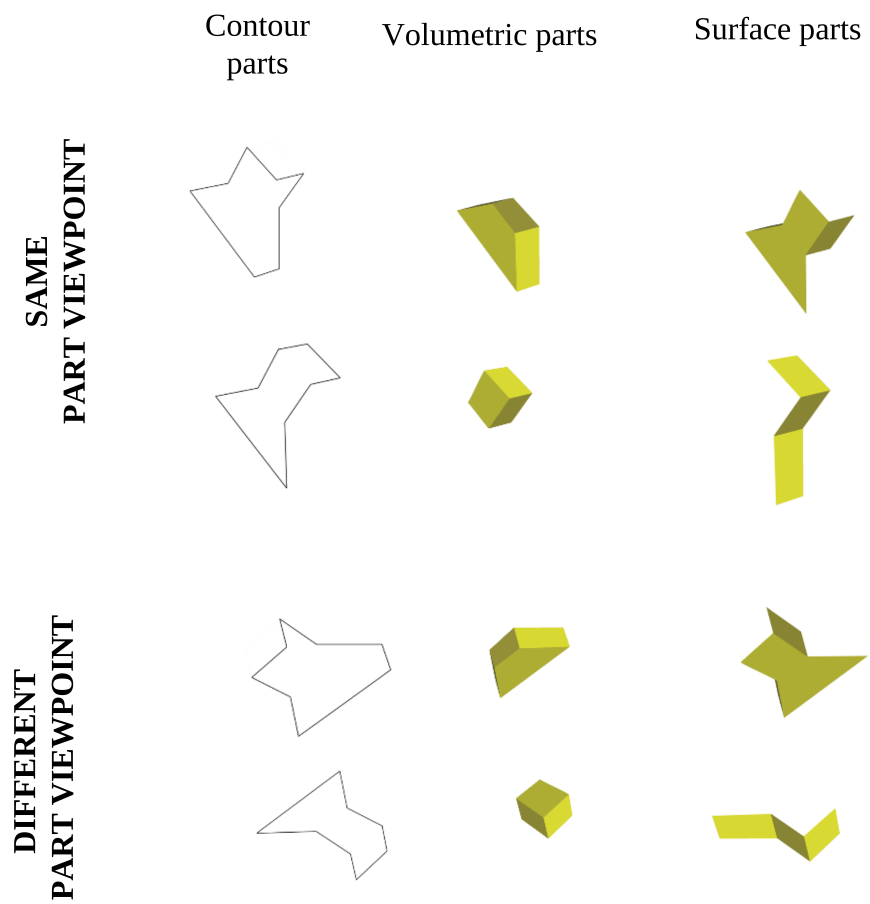
Contour parts
Volumetric parts
Surface parts
SAME PART VIEWPOINT
DIFFERENT PART VIEWPOINT

FIGURE 3

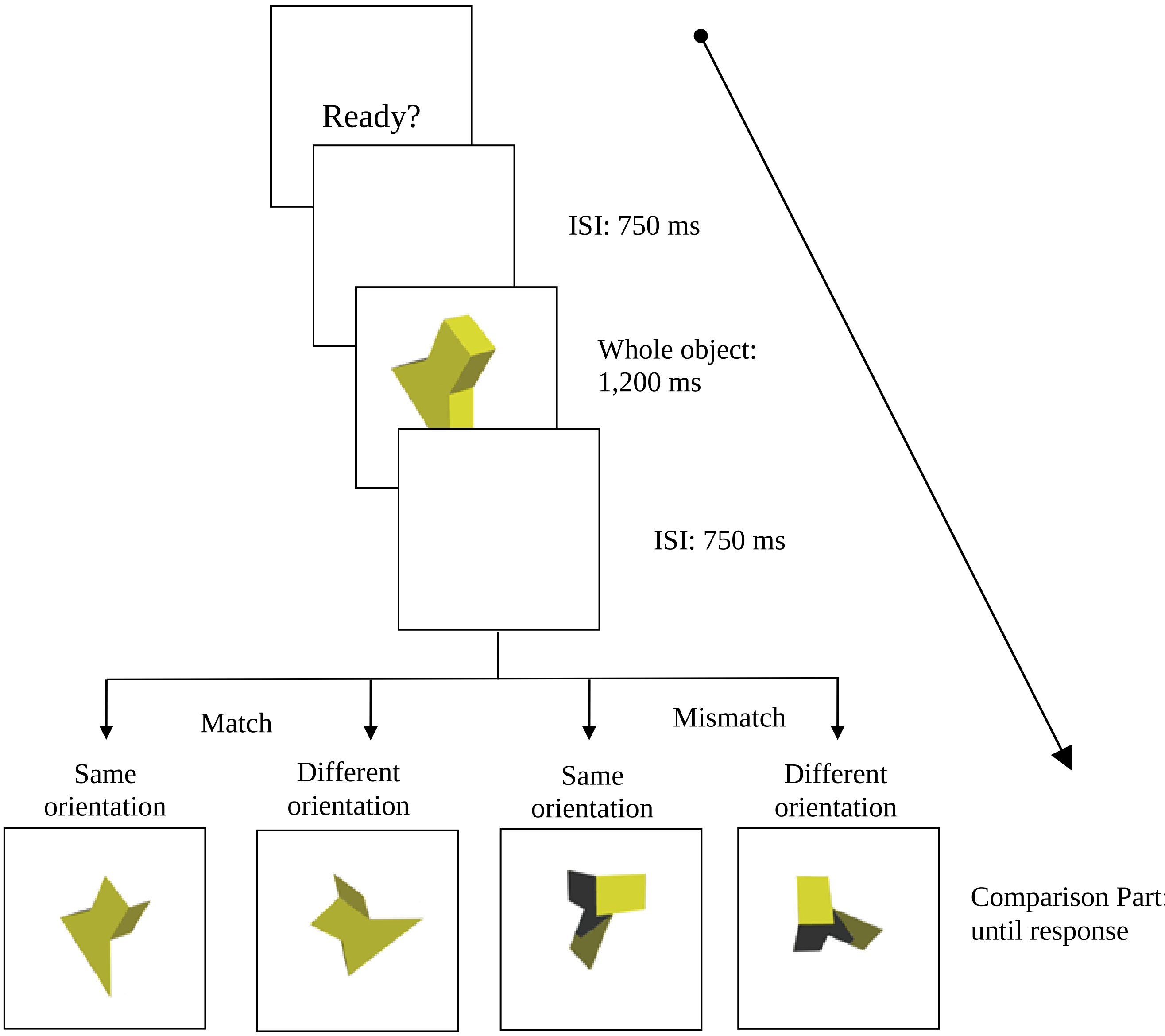
Ready?
ISI: 750 ms
Whole object:
1,200 ms
ISI: 750 ms
Match
Mismatch
Same
orientation
Different
orientation
Same
orientation
Different
orientation
Comparison Part:
until response

FIGURE 4A

Mean *d'* scores per Part Viewpoint, Part Type and Part Identity (collapsed across Group). The graphs illustrate the similarity in the pattern of matching (despite changes in part viewpoint relative to the whole object.

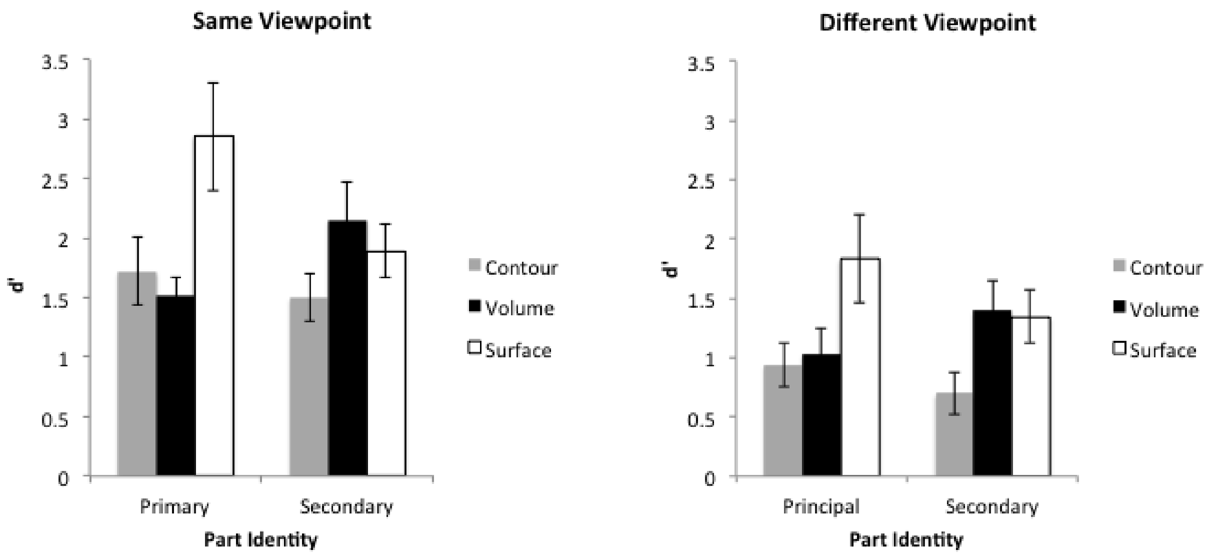

FIGURE 4B

Mean *d'* scores per Group and per Part Type for secondary parts only (see text for details). The graphs illustrate the similarity in the pattern of matching across contour, volumetric and surface parts in the two participant Groups.